Guest Editorial: Graph-Powered Machine Learning in Future-Generation Computing Systems

Shirui Pan[a], Shaoxiong Ji[b], Di Jin[c], Feng Xia[d] and Philip S. Yu[e]

[a]Monash University, Australia
[b]Aalto University, Finland
[c]Tianjin University, China
[d]Federation University Australia
[e]University of Illinois at Chicago, USA

ABSTRACT

1. Introduction

Recent years have witnessed a dramatic increase in graph applications due to advancements in information and communication technologies. In various applications, such as social networks, communication networks, the internet of things (IOTs), and human disease networks, graph data contains rich information and exhibits diverse characteristics. Specifically, graph data may come with the node or edge attributes showing the property of an entity or a connection, arise with signed or unsigned edges indicating the positive or negative relationships, form homogeneous or heterogeneous information networks modeling different scenarios and settings. Furthermore, in these applications, graph data is evolving and expanding more and more dynamically. The diverse, dynamic, and large-scale nature of graph data requires different data mining techniques and advanced machine learning methods. Meanwhile, the computing system evolves rapidly and becomes large-scale, collaborative, and distributed, with many computing principles proposed such as cloud computing, edge computing, and federated learning [6]. Learning from big graph data in future-generation computing systems considers the effectiveness of graph learning [22], scalability of large-scale computing, privacy-preserving under the federated computing setting with multi-source graphs, and graph dynamics in the distributed environment. Today's researchers have realized that novel graph learning theory, big graph-specific platforms, and advanced graph processing techniques are needed. Therefore, a set of research topics such as knowledge graph reasoning [5], graph self-supervised learning [12], temporal graph modeling [7], and graph embedding techniques [15, 21] have emerged, and applications such as graph-based anomaly detection [11, 13], community detection [8], social recommendation, social influence analytics are becoming important issues for the research community.

2. Contents of the special issue

This special issue focused on advanced data mining and machine learning methods and big graph data analytics applications in future generation computing systems. We received 62 valid paper submissions for this special issue. After several rounds of rigorous reviews and revisions, we decided to publish 15 of them.

The first article entitled "Event prediction based on evolutionary event ontology knowledge" [14] opens the special issue with an evolutionary event ontology knowledge base which provides a general-purpose ontology knowledge and facilities event extraction. The authors explore a framework with a pipeline for event extraction, evolutionary event recognition, and event prediction. The effectiveness of the proposed model is compared with other alternative methods.

Next, "Discovering communities from disjoint complex networks using multi-layer ant colony optimization" [4] proposes the Multi-Layer Ant Colony Optimization algorithm to detect communities in complex networks. The proposed method considers both Ratio Cut and Kernel K-means for the optimization to get the solution. Evaluated on both small and large-scale networks, the algorithm shows good performance in terms of normalized mutual information and modularity.

The article "Selection strategy in graph-based spreading dynamics with limited capacity" [23] studies effective strategies of altering epidemic spreading with limited capacity in social networks. The authors propose a graph-based diffusion model in which spreaders only contact a finite number of neighbors. Analytical and simulation results in artificial graphs prove that selection strategies change the final diffusion extent but do not alter the spreading threshold.

In "Mutual teaching for graph convolutional networks" [25], a new training strategy for graph neural networks (GCNs) is proposed. By using dual GCN models for training and enhancing each other, the performance of GCN is improved. The pseudo labels are exploited to calculate two new loss terms in the proposed algorithm.

The work "EagleMine: Vision-guided Micro-clusters recognition and collective anomaly detection" [3] performs anomaly detection in large-scale graphs with a vision-guided algorithm. The proposed algorithm, named EagleMine, hierarchically discovers node groups by utilizing a water-level tree with multiple resolutions according to the rule of visual recognition. It also identifies anomalous micro-clusters

ORCID(s):

which deviate from the majority nodes in a graph, achieving effective summarization performance compared to baselines.

The research “Recurrent-DC: A deep representation clustering model for university profiling based on academic graph" [9] proposes to perform clustering on academic graphs with a deep representation learning algorithm. The paper proposes a University Profiling Framework from the production and complexity point of view, based on the academic research which naturally forms heterogeneous graphs. Deep representation learning together with a Stacked Autoencoder are exploited for this task.

In “Graph-CAT: Graph Co-Attention Networks via local and global attribute augmentations" [24], the authors analyze the propagation strategies in two milestone methods, GCN and Graph Attention Network, to reveal their underlying philosophies. Based on the analysis, the authors propose the Graph Co-Attention Network method, which utilizes attention mechanisms on both the local and global attribute augmentations to improve the graph representation learning.

The work “Large-scale online multi-view graph neural network and applications" [10] proposes an attention-based Heterogeneous Multi-view Graph Neural Network for large scale graph representation learning. The proposed method overcomes the two limitations of existing studies: 1) most existing studies are transductive and mainly concentrate on homogeneous networks 2) most existing works are difficult to handle the real-time changing network structures. The proposed algorithm is evaluated on large-scale spam detection and link prediction tasks.

In the article “Deep spatial-temporal sequence modeling for multi-step passenger demand prediction" [2], authors propose an end-to-end deep learning framework for multi-step passenger demand prediction. The proposed algorithm considers the influence of both historical demand and heterogeneous external data and exploits both GCN and the long short-term memory algorithm to capture spatiotemporal relations.

The work “Influence maximization in social graphs based on community structure and node coverage gain" [20] proposes an influence maximization approach based on overlapping community structure and node coverage gain. The proposed algorithm first partitions the social graphs into overlapping groups and then uses a node coverage gain sensitive centrality measure to evaluate the influence of each node locally within its belonging group. In the last step, seed nodes are directly selected by combining the detected community structure.

The article “ACSIMCD: A 2-phase framework for detecting meaningful communities in dynamic social networks" [1] presents a 2-phase framework for community detection in dynamic social networks. Both content and structure information of nodes are utilized, and the proposed algorithm captures users’ interest with statistical and semantic measures. The dynamicity of social networks is analyzed by tracking users’ interests and behaviors. The detected communities can help identify users’ interests.

In “DGSD: Distributed graph representation via graph statistical properties" [17], authors propose a graph embedding method for centralized and parallel computing. This algorithm finds nodes’ local proximity by considering only nodes’ degree, common neighbors, and direct connectivity that allows it to run in the distributed environment. The proposed algorithm has linear space complexity for large graphs.

The article “Graph convolutional networks for graphs containing missing features" [18] proposed a new method for training graph convolutional networks where missing features are present in the graph data. The proposed algorithm integrates the processing of missing features and graph learning within the same neural network architecture in an end-to-end framework while maintaining similar computational complexity as GCN.

The work “Efficient search over incomplete knowledge graphs in binarized embedding space" [19] proposes to encode incomplete knowledge graphs (KGs) and graph queries in a Hamming space and presents a learning-to-hash model to learn binary embeddings for KG queries and entities. The hashed embedding can be used to discover target entities from incomplete KGs whilst the efficiency has been greatly improved.

Finally, the work “Detecting covert communities in multi-layer networks: A network embedding approach" [16] proposes a network embedding technique to find covert communities in multi-layer dark networks using a Log-BiLinear approach. Different from traditional approaches, the proposed method learns structural representations of nodes and relations simultaneously by capturing the position of a given node within a set of neighboring anchor-set. To identify the clusters (communities), clustering centroids are also learned as the representations of nodes and relations are extracted. This method is evaluated on real-world terrorist datasets to show its effectiveness.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the Editor-in-Chief, Prof Michela Taufer, and Special Content Editor, Prof Erik Cambria of Future Generation Computer Systems, for their valuable support during the preparation of this special issue. In addition, we are deeply grateful to the reviewers for their contributions to the decision process.

References

[1] Akachar, E., Ouhbi, B., Frikh, B., 2021. Acsimcd: A 2-phase framework for detecting meaningful communities in dynamic social networks. Future Generation Computer Systems 125, 399–420.

[2] Bai, L., Yao, L., Wang, X., Li, C., Zhang, X., 2021. Deep spatial–temporal sequence modeling for multi-step passenger demand prediction. Future Generation Computer Systems 121, 25–34.

[3] Feng, W., Liu, S., Faloutsos, C., Hooi, B., Shen, H., Cheng, X., 2021. Eaglemine: Vision-guided micro-clusters recognition and collective anomaly detection. Future Generation Computer Systems 115, 236–250.

[4] Imtiaz, Z.B., Manzoor, A., ul Islam, S., Judge, M.A., Choo, K.K.R., Rodrigues, J.J., 2021. Discovering communities from disjoint complex networks using multi-layer ant colony optimization. Future Generation Computer Systems 115, 659–670.

[5] Ji, S., Pan, S., Cambria, E., Marttinen, P., Philip, S.Y., 2021a. A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems .

[6] Ji, S., Saravirta, T., Pan, S., Long, G., Walid, A., 2021b. Emerging trends in federated learning: From model fusion to federated x learning. arXiv preprint arXiv:2102.12920 .

[7] Jiao, P., Guo, X., Jing, X., He, D., Wu, H., Pan, S., Gong, M., Wang, W., 2021. Temporal network embedding for link prediction via vae joint attention mechanism. IEEE Transactions on Neural Networks and Learning Systems .

[8] Jin, D., Yu, Z., Jiao, P., Pan, S., Yu, P.S., Zhang, W., 2021. A survey of community detection approaches: From statistical modeling to deep learning. arXiv preprint arXiv:2101.01669 .

[9] Kong, X., Li, J., Wang, L., Shen, G., Sun, Y., Lee, I., 2021. Recurrent-dc: A deep representation clustering model for university profiling based on academic graph. Future Generation Computer Systems 116, 156–167.

[10] Li, Z., Xing, Y., Huang, J., Wang, H., Gao, J., Yu, G., 2021. Large-scale online multi-view graph neural network and applications. Future Generation Computer Systems 116, 145–155.

[11] Liu, Y., Li, Z., Pan, S., Gong, C., Zhou, C., Karypis, G., 2021a. Anomaly detection on attributed networks via contrastive self-supervised learning. IEEE Transactions on Neural Networks and Learning Systems .

[12] Liu, Y., Pan, S., Jin, M., Zhou, C., Xia, F., Yu, P.S., 2021b. Graph self-supervised learning: A survey. arXiv preprint arXiv:2103.00111 .

[13] Liu, Y., Pan, S., Wang, Y.G., Xiong, F., Wang, L., Lee, V., 2021c. Anomaly detection in dynamic graphs via transformer. arXiv preprint arXiv:2106.09876 .

[14] Mao, Q., Li, X., Peng, H., Li, J., He, D., Guo, S., He, M., Wang, L., 2021. Event prediction based on evolutionary event ontology knowledge. Future Generation Computer Systems 115, 76–89.

[15] Pan, S., Hu, R., Fung, S.f., Long, G., Jiang, J., Zhang, C., 2019. Learning graph embedding with adversarial training methods. IEEE transactions on cybernetics 50, 2475–2487.

[16] Pourhabibi, T., Ong, K.L., Boo, Y.L., Kam, B.H., 2021. Detecting covert communities in multi-layer networks: A network embedding approach. Future Generation Computer Systems .

[17] Said, A., Hassan, S.U., Tuarob, S., Nawaz, R., Shabbir, M., 2021. Dgsd: Distributed graph representation via graph statistical properties. Future Generation Computer Systems 119, 166–175.

[18] Taguchi, H., Liu, X., Murata, T., 2021. Graph convolutional networks for graphs containing missing features. Future Generation Computer Systems 117, 155–168.

[19] Wang, M., Chen, W., Wang, S., Jiang, Y., Yao, L., Qi, G., 2021a. Efficient search over incomplete knowledge graphs in binarized embedding space. Future Generation Computer Systems 123, 24–34.

[20] Wang, Z., Sun, C., Xi, J., Li, X., 2021b. Influence maximization in social graphs based on community structure and node coverage gain. Future Generation Computer Systems 118, 327–338.

[21] Wu, M., Pan, S., Du, L., Zhu, X., 2021. Learning graph neural networks with positive and unlabeled nodes. ACM Transactions on Knowledge Discovery from Data .

[22] Xia, F., Sun, K., Yu, S., Aziz, A., Wan, L., Pan, S., Liu, H., 2021. Graph learning: A survey. IEEE Transactions on Artificial Intelligence 1, 1–1.

[23] Xiong, F., Zheng, Y., Ding, W., Wang, H., Wang, X., Chen, H., 2021. Selection strategy in graph-based spreading dynamics with limited capacity. Future Generation Computer Systems 114, 307–317.

[24] Yang, L., Li, W., Guo, Y., Gu, J., 2021. Graph-cat: Graph co-attention networks via local and global attribute augmentations. Future Generation Computer Systems 118, 170–179.

[25] Zhan, K., Niu, C., 2021. Mutual teaching for graph convolutional networks. Future Generation Computer Systems 115, 837–843.